

Advancing Interdisciplinary Science with Foundation Models

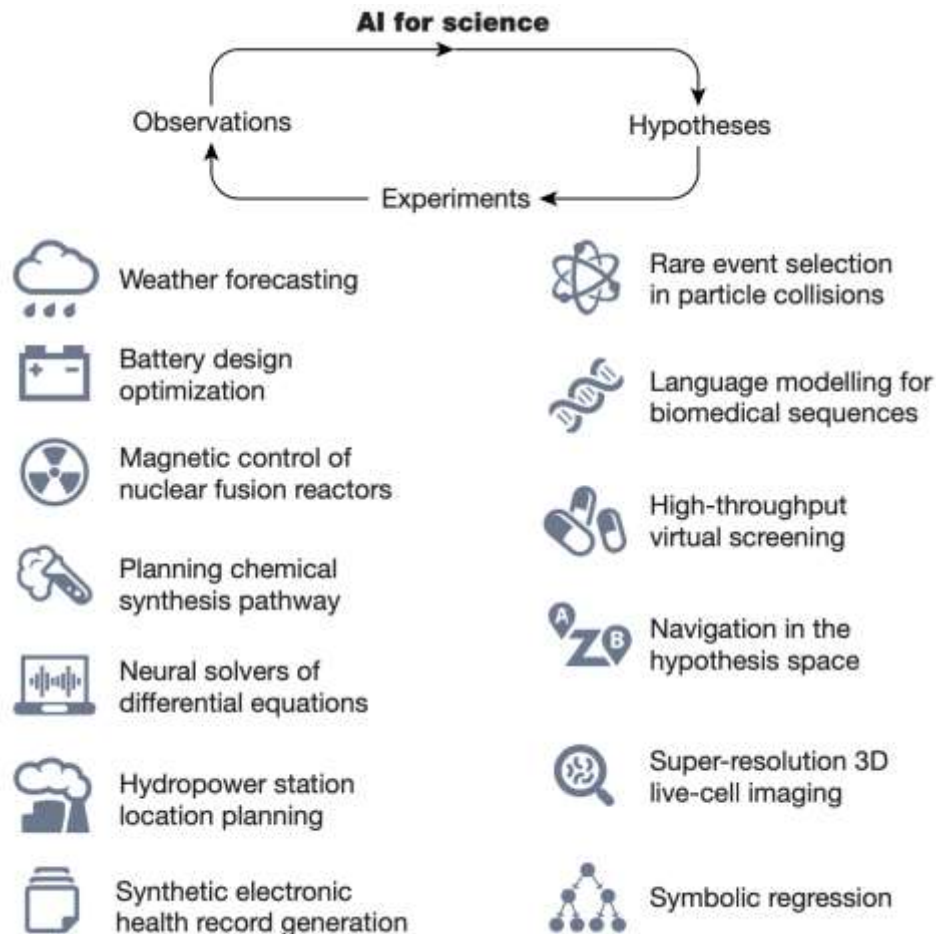
Sameera Horawalavithana, Ph.D.
yasanka.horawalavithana@pnnl.gov



Outline

- AI for Science
- Multimodal Scientific Data
- Building Foundation Models for Science
 - ATLANTIC: Structure-Aware LLMs for Interdisciplinary Science
 - SCITUNE: Multimodal Agents for Scientific Reasoning
- Key Takeaways

AI for Science

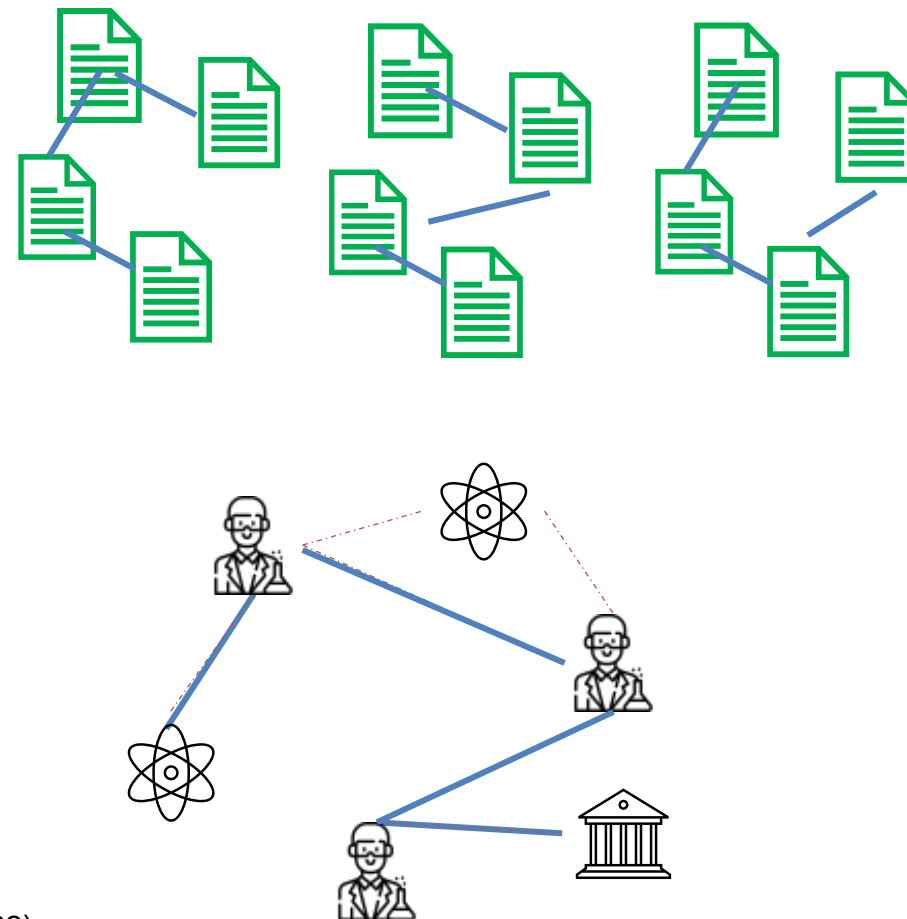
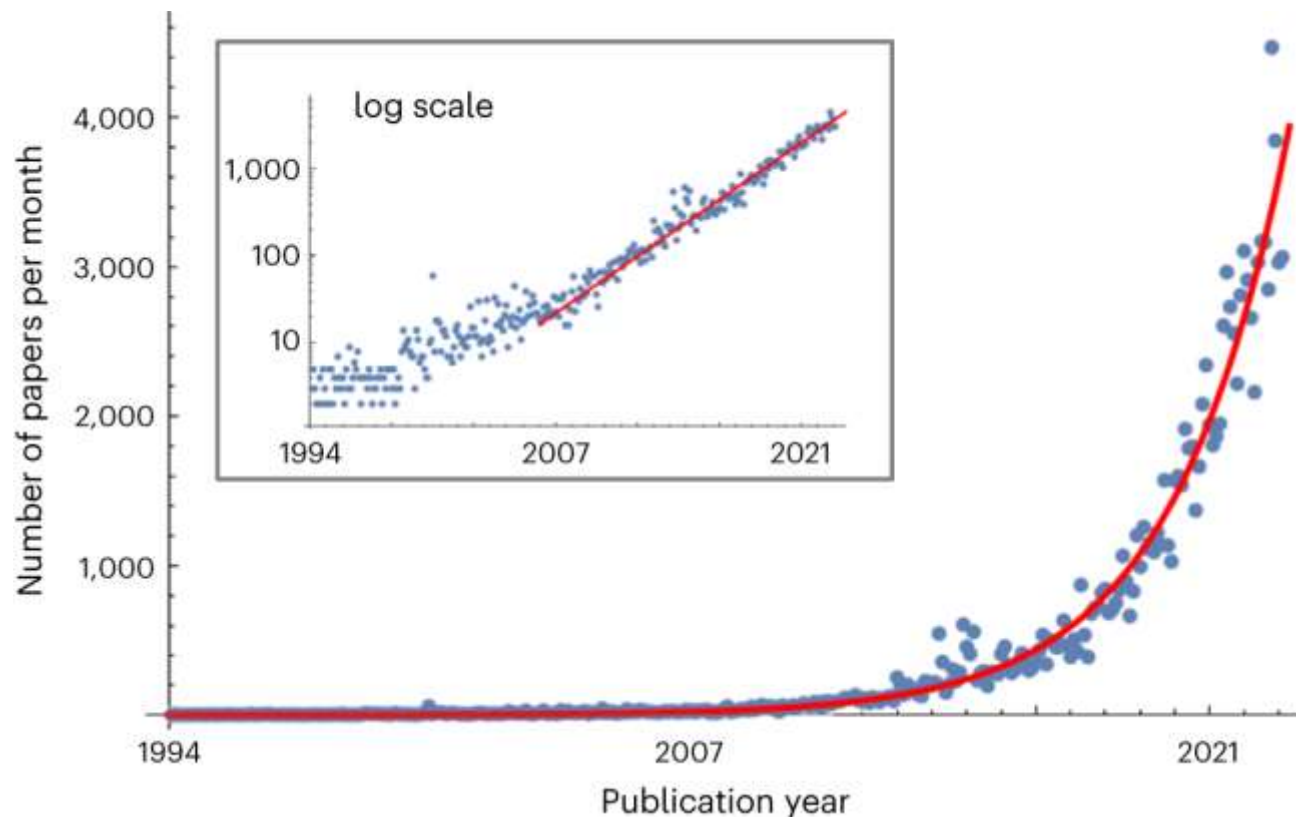


Wang, Hanchen, et al. "Scientific discovery in the age of artificial intelligence." *Nature* 620.7972 (2023): 47-60.



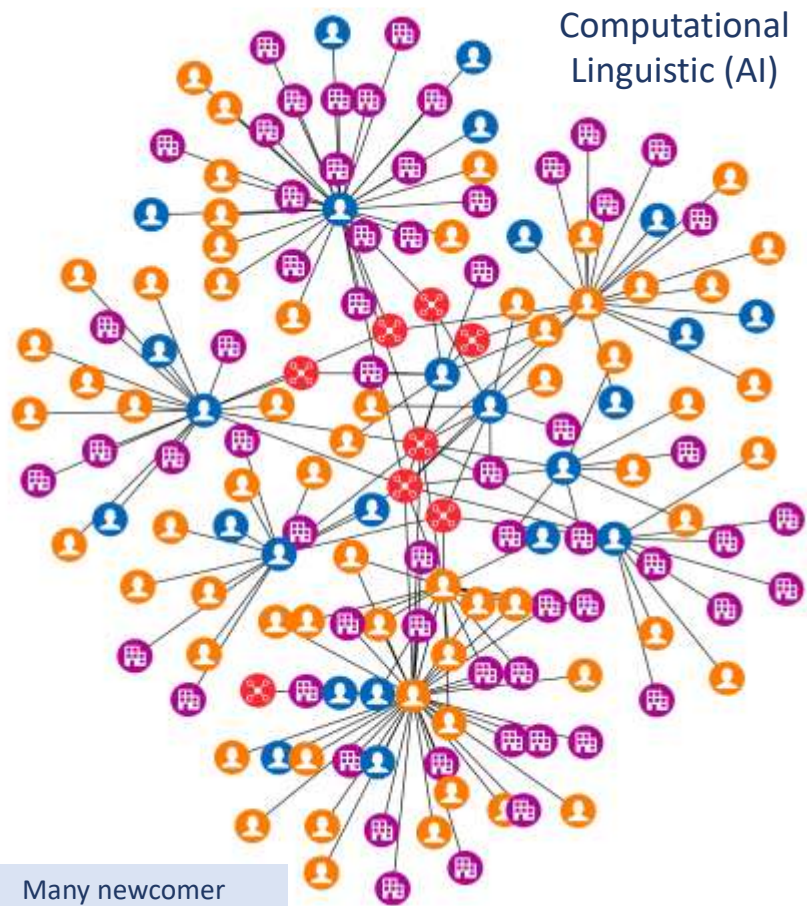
The Evolving Scientific Community

The doubling rate of papers per month is roughly 23 months.

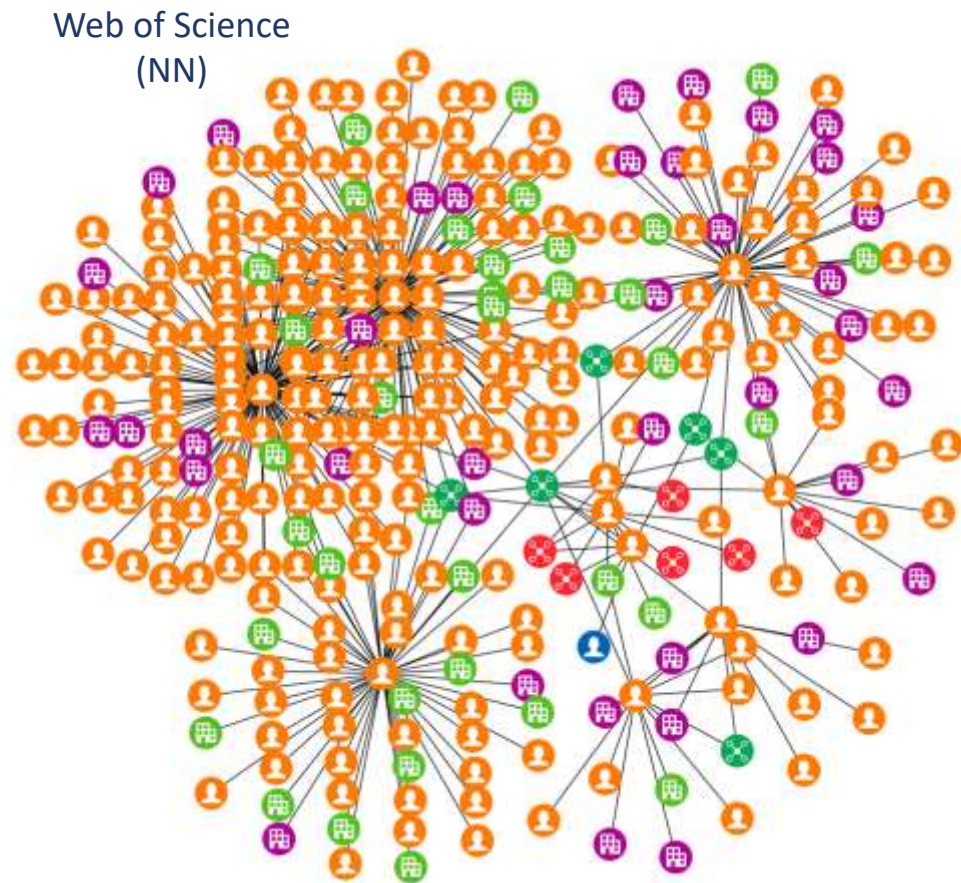


Krenn, M., Buffoni, L., Coutinho, B., Eppel, S., Foster, J. G., Gritsevskiy, A., ... & Kopp, M. (2023). Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. *Nature Machine Intelligence*, 1-10.

The Evolving Scientific Community



Many newcomer scientists with new institution partnerships



Many incumbent scientists with repeated institution partnerships



Machine Learning (AI)

Multimodal Scientific Data for AI Models

Tables



Journal of Materials Chemistry A

View Article Online
Paper

Table 1 The photovoltaic parameters of the (BA/GA)₂MA₂Pb₂I₈ based PSCs without and with formamide addition on different substrates*

			V_{oc} [V]	FF	J_{sc} [mA cm ⁻²]	PCP _{average} [%]	IR [%]	J_{tot} [mA cm ⁻²]
DMF w/o f	PEDOT:PSS	Reverse	1.09	0.79	12.23	15.36/16.91	4.6	13.12
		Forward	0.96	0.76	12.03	15.07		
DMF w/ f	PEDOT:PSS	Reverse	1.00	0.77	17.03	13.23/12.80	4.5	16.30
		Forward	1.00	0.76	16.48	13.92		
DMF w/ f	PCP-Ia	Reverse	1.17	0.81	16.99	15.16/15.19	4.2	16.98
		Forward	1.16	0.77	17.00	15.19		

* The average PCPs were calculated based on 40 devices.

Equations



Text



Fig. 2 Cross-sectional SEM images of the (BA/GA)₂MA₂Pb₂I₈ perovskite films without and with formamide addition on different substrates.

Single X-ray scattering (GIWAXS) patterns of (BA/GA)₂MA₂Pb₂I₈ perovskite films with and without formamide were recorded (Fig. 3a and b). The Bragg spots for the perovskite film w/o f span a slightly wider angle than the film w/ f, indicating that the formamide addition improves the growth orientation of the crystals in the perovskite film. The corresponding X-ray diffraction (XRD) patterns and parameters are presented in Fig. 3c and Table 2. The diffraction peaks at 14.1° and 16.3° can be attributed to the (111) and (202) lattice planes of the perovskite films, respectively.^{24,26} For the film with 3% formamide addition, the intensity enhancement of the (111) and (202)

peaks can be detected. FWHM values of (111) and (202) peaks of the perovskite films narrow down from 0.216° and 0.252° (without formamide addition) to 0.209° and 0.227° (with formamide addition), indicating the enhancement of grain size, which is in favor of improving carrier transport and suppressing charge recombination.^{44,45}

The effect on the carrier lifetime of the (BA/GA)₂MA₂Pb₂I₈ perovskite films deposited from DMF with and without formamide on quartz substrates was evaluated through time-resolved PL decay spectra (TRPL), as shown in Fig. 4a. The corresponding calculated results are summarized in Table S1†. The TRPL curve was fitted with a bi-exponential decay function, $I(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2)$, assuming a fast decay and a slow decay process.⁴⁶ The recombination lifetime, calculated by the formula of $\tau_{eff} = 1/\tau_1 + \tau_2$, is $\tau_{eff} = \tau_1 \tau_2 / (\tau_1 + \tau_2)$. τ_{eff} of A₁ is 1.0 ns, τ_{eff} of A₂ is 1.0 ns, τ_{eff} of A₃ is 1.0 ns, τ_{eff} of A₄ is 1.0 ns, τ_{eff} of A₅ is 1.0 ns, τ_{eff} of A₆ is 1.0 ns, τ_{eff} of A₇ is 1.0 ns, τ_{eff} of A₈ is 1.0 ns, τ_{eff} of A₉ is 1.0 ns, τ_{eff} of A₁₀ is 1.0 ns, τ_{eff} of A₁₁ is 1.0 ns, τ_{eff} of A₁₂ is 1.0 ns, τ_{eff} of A₁₃ is 1.0 ns, τ_{eff} of A₁₄ is 1.0 ns, τ_{eff} of A₁₅ is 1.0 ns, τ_{eff} of A₁₆ is 1.0 ns, τ_{eff} of A₁₇ is 1.0 ns, τ_{eff} of A₁₈ is 1.0 ns, τ_{eff} of A₁₉ is 1.0 ns, τ_{eff} of A₂₀ is 1.0 ns, τ_{eff} of A₂₁ is 1.0 ns, τ_{eff} of A₂₂ is 1.0 ns, τ_{eff} of A₂₃ is 1.0 ns, τ_{eff} of A₂₄ is 1.0 ns, τ_{eff} of A₂₅ is 1.0 ns, τ_{eff} of A₂₆ is 1.0 ns, τ_{eff} of A₂₇ is 1.0 ns, τ_{eff} of A₂₈ is 1.0 ns, τ_{eff} of A₂₉ is 1.0 ns, τ_{eff} of A₃₀ is 1.0 ns, τ_{eff} of A₃₁ is 1.0 ns, τ_{eff} of A₃₂ is 1.0 ns, τ_{eff} of A₃₃ is 1.0 ns, τ_{eff} of A₃₄ is 1.0 ns, τ_{eff} of A₃₅ is 1.0 ns, τ_{eff} of A₃₆ is 1.0 ns, τ_{eff} of A₃₇ is 1.0 ns, τ_{eff} of A₃₈ is 1.0 ns, τ_{eff} of A₃₉ is 1.0 ns, τ_{eff} of A₄₀ is 1.0 ns, τ_{eff} of A₄₁ is 1.0 ns, τ_{eff} of A₄₂ is 1.0 ns, τ_{eff} of A₄₃ is 1.0 ns, τ_{eff} of A₄₄ is 1.0 ns, τ_{eff} of A₄₅ is 1.0 ns, τ_{eff} of A₄₆ is 1.0 ns, τ_{eff} of A₄₇ is 1.0 ns, τ_{eff} of A₄₈ is 1.0 ns, τ_{eff} of A₄₉ is 1.0 ns, τ_{eff} of A₅₀ is 1.0 ns, τ_{eff} of A₅₁ is 1.0 ns, τ_{eff} of A₅₂ is 1.0 ns, τ_{eff} of A₅₃ is 1.0 ns, τ_{eff} of A₅₄ is 1.0 ns, τ_{eff} of A₅₅ is 1.0 ns, τ_{eff} of A₅₆ is 1.0 ns, τ_{eff} of A₅₇ is 1.0 ns, τ_{eff} of A₅₈ is 1.0 ns, τ_{eff} of A₅₉ is 1.0 ns, τ_{eff} of A₆₀ is 1.0 ns, τ_{eff} of A₆₁ is 1.0 ns, τ_{eff} of A₆₂ is 1.0 ns, τ_{eff} of A₆₃ is 1.0 ns, τ_{eff} of A₆₄ is 1.0 ns, τ_{eff} of A₆₅ is 1.0 ns, τ_{eff} of A₆₆ is 1.0 ns, τ_{eff} of A₆₇ is 1.0 ns, τ_{eff} of A₆₈ is 1.0 ns, τ_{eff} of A₆₉ is 1.0 ns, τ_{eff} of A₇₀ is 1.0 ns, τ_{eff} of A₇₁ is 1.0 ns, τ_{eff} of A₇₂ is 1.0 ns, τ_{eff} of A₇₃ is 1.0 ns, τ_{eff} of A₇₄ is 1.0 ns, τ_{eff} of A₇₅ is 1.0 ns, τ_{eff} of A₇₆ is 1.0 ns, τ_{eff} of A₇₇ is 1.0 ns, τ_{eff} of A₇₈ is 1.0 ns, τ_{eff} of A₇₉ is 1.0 ns, τ_{eff} of A₈₀ is 1.0 ns, τ_{eff} of A₈₁ is 1.0 ns, τ_{eff} of A₈₂ is 1.0 ns, τ_{eff} of A₈₃ is 1.0 ns, τ_{eff} of A₈₄ is 1.0 ns, τ_{eff} of A₈₅ is 1.0 ns, τ_{eff} of A₈₆ is 1.0 ns, τ_{eff} of A₈₇ is 1.0 ns, τ_{eff} of A₈₈ is 1.0 ns, τ_{eff} of A₈₉ is 1.0 ns, τ_{eff} of A₉₀ is 1.0 ns, τ_{eff} of A₉₁ is 1.0 ns, τ_{eff} of A₉₂ is 1.0 ns, τ_{eff} of A₉₃ is 1.0 ns, τ_{eff} of A₉₄ is 1.0 ns, τ_{eff} of A₉₅ is 1.0 ns, τ_{eff} of A₉₆ is 1.0 ns, τ_{eff} of A₉₇ is 1.0 ns, τ_{eff} of A₉₈ is 1.0 ns, τ_{eff} of A₉₉ is 1.0 ns, τ_{eff} of A₁₀₀ is 1.0 ns, τ_{eff} of A₁₀₁ is 1.0 ns, τ_{eff} of A₁₀₂ is 1.0 ns, τ_{eff} of A₁₀₃ is 1.0 ns, τ_{eff} of A₁₀₄ is 1.0 ns, τ_{eff} of A₁₀₅ is 1.0 ns, τ_{eff} of A₁₀₆ is 1.0 ns, τ_{eff} of A₁₀₇ is 1.0 ns, τ_{eff} of A₁₀₈ is 1.0 ns, τ_{eff} of A₁₀₉ is 1.0 ns, τ_{eff} of A₁₁₀ is 1.0 ns, τ_{eff} of A₁₁₁ is 1.0 ns, τ_{eff} of A₁₁₂ is 1.0 ns, τ_{eff} of A₁₁₃ is 1.0 ns, τ_{eff} of A₁₁₄ is 1.0 ns, τ_{eff} of A₁₁₅ is 1.0 ns, τ_{eff} of A₁₁₆ is 1.0 ns, τ_{eff} of A₁₁₇ is 1.0 ns, τ_{eff} of A₁₁₈ is 1.0 ns, τ_{eff} of A₁₁₉ is 1.0 ns, τ_{eff} of A₁₂₀ is 1.0 ns, τ_{eff} of A₁₂₁ is 1.0 ns, τ_{eff} of A₁₂₂ is 1.0 ns, τ_{eff} of A₁₂₃ is 1.0 ns, τ_{eff} of A₁₂₄ is 1.0 ns, τ_{eff} of A₁₂₅ is 1.0 ns, τ_{eff} of A₁₂₆ is 1.0 ns, τ_{eff} of A₁₂₇ is 1.0 ns, τ_{eff} of A₁₂₈ is 1.0 ns, τ_{eff} of A₁₂₉ is 1.0 ns, τ_{eff} of A₁₃₀ is 1.0 ns, τ_{eff} of A₁₃₁ is 1.0 ns, τ_{eff} of A₁₃₂ is 1.0 ns, τ_{eff} of A₁₃₃ is 1.0 ns, τ_{eff} of A₁₃₄ is 1.0 ns, τ_{eff} of A₁₃₅ is 1.0 ns, τ_{eff} of A₁₃₆ is 1.0 ns, τ_{eff} of A₁₃₇ is 1.0 ns, τ_{eff} of A₁₃₈ is 1.0 ns, τ_{eff} of A₁₃₉ is 1.0 ns, τ_{eff} of A₁₄₀ is 1.0 ns, τ_{eff} of A₁₄₁ is 1.0 ns, τ_{eff} of A₁₄₂ is 1.0 ns, τ_{eff} of A₁₄₃ is 1.0 ns, τ_{eff} of A₁₄₄ is 1.0 ns, τ_{eff} of A₁₄₅ is 1.0 ns, τ_{eff} of A₁₄₆ is 1.0 ns, τ_{eff} of A₁₄₇ is 1.0 ns, τ_{eff} of A₁₄₈ is 1.0 ns, τ_{eff} of A₁₄₉ is 1.0 ns, τ_{eff} of A₁₅₀ is 1.0 ns, τ_{eff} of A₁₅₁ is 1.0 ns, τ_{eff} of A₁₅₂ is 1.0 ns, τ_{eff} of A₁₅₃ is 1.0 ns, τ_{eff} of A₁₅₄ is 1.0 ns, τ_{eff} of A₁₅₅ is 1.0 ns, τ_{eff} of A₁₅₆ is 1.0 ns, τ_{eff} of A₁₅₇ is 1.0 ns, τ_{eff} of A₁₅₈ is 1.0 ns, τ_{eff} of A₁₅₉ is 1.0 ns, τ_{eff} of A₁₆₀ is 1.0 ns, τ_{eff} of A₁₆₁ is 1.0 ns, τ_{eff} of A₁₆₂ is 1.0 ns, τ_{eff} of A₁₆₃ is 1.0 ns, τ_{eff} of A₁₆₄ is 1.0 ns, τ_{eff} of A₁₆₅ is 1.0 ns, τ_{eff} of A₁₆₆ is 1.0 ns, τ_{eff} of A₁₆₇ is 1.0 ns, τ_{eff} of A₁₆₈ is 1.0 ns, τ_{eff} of A₁₆₉ is 1.0 ns, τ_{eff} of A₁₇₀ is 1.0 ns, τ_{eff} of A₁₇₁ is 1.0 ns, τ_{eff} of A₁₇₂ is 1.0 ns, τ_{eff} of A₁₇₃ is 1.0 ns, τ_{eff} of A₁₇₄ is 1.0 ns, τ_{eff} of A₁₇₅ is 1.0 ns, τ_{eff} of A₁₇₆ is 1.0 ns, τ_{eff} of A₁₇₇ is 1.0 ns, τ_{eff} of A₁₇₈ is 1.0 ns, τ_{eff} of A₁₇₉ is 1.0 ns, τ_{eff} of A₁₈₀ is 1.0 ns, τ_{eff} of A₁₈₁ is 1.0 ns, τ_{eff} of A₁₈₂ is 1.0 ns, τ_{eff} of A₁₈₃ is 1.0 ns, τ_{eff} of A₁₈₄ is 1.0 ns, τ_{eff} of A₁₈₅ is 1.0 ns, τ_{eff} of A₁₈₆ is 1.0 ns, τ_{eff} of A₁₈₇ is 1.0 ns, τ_{eff} of A₁₈₈ is 1.0 ns, τ_{eff} of A₁₈₉ is 1.0 ns, τ_{eff} of A₁₉₀ is 1.0 ns, τ_{eff} of A₁₉₁ is 1.0 ns, τ_{eff} of A₁₉₂ is 1.0 ns, τ_{eff} of A₁₉₃ is 1.0 ns, τ_{eff} of A₁₉₄ is 1.0 ns, τ_{eff} of A₁₉₅ is 1.0 ns, τ_{eff} of A₁₉₆ is 1.0 ns, τ_{eff} of A₁₉₇ is 1.0 ns, τ_{eff} of A₁₉₈ is 1.0 ns, τ_{eff} of A₁₉₉ is 1.0 ns, τ_{eff} of A₂₀₀ is 1.0 ns, τ_{eff} of A₂₀₁ is 1.0 ns, τ_{eff} of A₂₀₂ is 1.0 ns, τ_{eff} of A₂₀₃ is 1.0 ns, τ_{eff} of A₂₀₄ is 1.0 ns, τ_{eff} of A₂₀₅ is 1.0 ns, τ_{eff} of A₂₀₆ is 1.0 ns, τ_{eff} of A₂₀₇ is 1.0 ns, τ_{eff} of A₂₀₈ is 1.0 ns, τ_{eff} of A₂₀₉ is 1.0 ns, τ_{eff} of A₂₁₀ is 1.0 ns, τ_{eff} of A₂₁₁ is 1.0 ns, τ_{eff} of A₂₁₂ is 1.0 ns, τ_{eff} of A₂₁₃ is 1.0 ns, τ_{eff} of A₂₁₄ is 1.0 ns, τ_{eff} of A₂₁₅ is 1.0 ns, τ_{eff} of A₂₁₆ is 1.0 ns, τ_{eff} of A₂₁₇ is 1.0 ns, τ_{eff} of A₂₁₈ is 1.0 ns, τ_{eff} of A₂₁₉ is 1.0 ns, τ_{eff} of A₂₂₀ is 1.0 ns, τ_{eff} of A₂₂₁ is 1.0 ns, τ_{eff} of A₂₂₂ is 1.0 ns, τ_{eff} of A₂₂₃ is 1.0 ns, τ_{eff} of A₂₂₄ is 1.0 ns, τ_{eff} of A₂₂₅ is 1.0 ns, τ_{eff} of A₂₂₆ is 1.0 ns, τ_{eff} of A₂₂₇ is 1.0 ns, τ_{eff} of A₂₂₈ is 1.0 ns, τ_{eff} of A₂₂₉ is 1.0 ns, τ_{eff} of A₂₃₀ is 1.0 ns, τ_{eff} of A₂₃₁ is 1.0 ns, τ_{eff} of A₂₃₂ is 1.0 ns, τ_{eff} of A₂₃₃ is 1.0 ns, τ_{eff} of A₂₃₄ is 1.0 ns, τ_{eff} of A₂₃₅ is 1.0 ns, τ_{eff} of A₂₃₆ is 1.0 ns, τ_{eff} of A₂₃₇ is 1.0 ns, τ_{eff} of A₂₃₈ is 1.0 ns, τ_{eff} of A₂₃₉ is 1.0 ns, τ_{eff} of A₂₄₀ is 1.0 ns, τ_{eff} of A₂₄₁ is 1.0 ns, τ_{eff} of A₂₄₂ is 1.0 ns, τ_{eff} of A₂₄₃ is 1.0 ns, τ_{eff} of A₂₄₄ is 1.0 ns, τ_{eff} of A₂₄₅ is 1.0 ns, τ_{eff} of A₂₄₆ is 1.0 ns, τ_{eff} of A₂₄₇ is 1.0 ns, τ_{eff} of A₂₄₈ is 1.0 ns, τ_{eff} of A₂₄₉ is 1.0 ns, τ_{eff} of A₂₅₀ is 1.0 ns, τ_{eff} of A₂₅₁ is 1.0 ns, τ_{eff} of A₂₅₂ is 1.0 ns, τ_{eff} of A₂₅₃ is 1.0 ns, τ_{eff} of A₂₅₄ is 1.0 ns, τ_{eff} of A₂₅₅ is 1.0 ns, τ_{eff} of A₂₅₆ is 1.0 ns, τ_{eff} of A₂₅₇ is 1.0 ns, τ_{eff} of A₂₅₈ is 1.0 ns, τ_{eff} of A₂₅₉ is 1.0 ns, τ_{eff} of A₂₆₀ is 1.0 ns, τ_{eff} of A₂₆₁ is 1.0 ns, τ_{eff} of A₂₆₂ is 1.0 ns, τ_{eff} of A₂₆₃ is 1.0 ns, τ_{eff} of A₂₆₄ is 1.0 ns, τ_{eff} of A₂₆₅ is 1.0 ns, τ_{eff} of A₂₆₆ is 1.0 ns, τ_{eff} of A₂₆₇ is 1.0 ns, τ_{eff} of A₂₆₈ is 1.0 ns, τ_{eff} of A₂₆₉ is 1.0 ns, τ_{eff} of A₂₇₀ is 1.0 ns, τ_{eff} of A₂₇₁ is 1.0 ns, τ_{eff} of A₂₇₂ is 1.0 ns, τ_{eff} of A₂₇₃ is 1.0 ns, τ_{eff} of A₂₇₄ is 1.0 ns, τ_{eff} of A₂₇₅ is 1.0 ns, τ_{eff} of A₂₇₆ is 1.0 ns, τ_{eff} of A₂₇₇ is 1.0 ns, τ_{eff} of A₂₇₈ is 1.0 ns, τ_{eff} of A₂₇₉ is 1.0 ns, τ_{eff} of A₂₈₀ is 1.0 ns, τ_{eff} of A₂₈₁ is 1.0 ns, τ_{eff} of A₂₈₂ is 1.0 ns, τ_{eff} of A₂₈₃ is 1.0 ns, τ_{eff} of A₂₈₄ is 1.0 ns, τ_{eff} of A₂₈₅ is 1.0 ns, τ_{eff} of A₂₈₆ is 1.0 ns, τ_{eff} of A₂₈₇ is 1.0 ns, τ_{eff} of A₂₈₈ is 1.0 ns, τ_{eff} of A₂₈₉ is 1.0 ns, τ_{eff} of A₂₉₀ is 1.0 ns, τ_{eff} of A₂₉₁ is 1.0 ns, τ_{eff} of A₂₉₂ is 1.0 ns, τ_{eff} of A₂₉₃ is 1.0 ns, τ_{eff} of A₂₉₄ is 1.0 ns, τ_{eff} of A₂₉₅ is 1.0 ns, τ_{eff} of A₂₉₆ is 1.0 ns, τ_{eff} of A₂₉₇ is 1.0 ns, τ_{eff} of A₂₉₈ is 1.0 ns, τ_{eff} of A₂₉₉ is 1.0 ns, τ_{eff} of A₃₀₀ is 1.0 ns, τ_{eff} of A₃₀₁ is 1.0 ns, τ_{eff} of A₃₀₂ is 1.0 ns, τ_{eff} of A₃₀₃ is 1.0 ns, τ_{eff} of A₃₀₄ is 1.0 ns, τ_{eff} of A₃₀₅ is 1.0 ns, τ_{eff} of A₃₀₆ is 1.0 ns, τ_{eff} of A₃₀₇ is 1.0 ns, τ_{eff} of A₃₀₈ is 1.0 ns, τ_{eff} of A₃₀₉ is 1.0 ns, τ_{eff} of A₃₁₀ is 1.0 ns, τ_{eff} of A₃₁₁ is 1.0 ns, τ_{eff} of A₃₁₂ is 1.0 ns, τ_{eff} of A₃₁₃ is 1.0 ns, τ_{eff} of A₃₁₄ is 1.0 ns, τ_{eff} of A₃₁₅ is 1.0 ns, τ_{eff} of A₃₁₆ is 1.0 ns, τ_{eff} of A₃₁₇ is 1.0 ns, τ_{eff} of A₃₁₈ is 1.0 ns, τ_{eff} of A₃₁₉ is 1.0 ns, τ_{eff} of A₃₂₀ is 1.0 ns, τ_{eff} of A₃₂₁ is 1.0 ns, τ_{eff} of A₃₂₂ is 1.0 ns, τ_{eff} of A₃₂₃ is 1.0 ns, τ_{eff} of A₃₂₄ is 1.0 ns, τ_{eff} of A₃₂₅ is 1.0 ns, τ_{eff} of A₃₂₆ is 1.0 ns, τ_{eff} of A₃₂₇ is 1.0 ns, τ_{eff} of A₃₂₈ is 1.0 ns, τ_{eff} of A₃₂₉ is 1.0 ns, τ_{eff} of A₃₃₀ is 1.0 ns, τ_{eff} of A₃₃₁ is 1.0 ns, τ_{eff} of A₃₃₂ is 1.0 ns, τ_{eff} of A₃₃₃ is 1.0 ns, τ_{eff} of A₃₃₄ is 1.0 ns, τ_{eff} of A₃₃₅ is 1.0 ns, τ_{eff} of A₃₃₆ is 1.0 ns, τ_{eff} of A₃₃₇ is 1.0 ns, τ_{eff} of A₃₃₈ is 1.0 ns, τ_{eff} of A₃₃₉ is 1.0 ns, τ_{eff} of A₃₄₀ is 1.0 ns, τ_{eff} of A₃₄₁ is 1.0 ns, τ_{eff} of A₃₄₂ is 1.0 ns, τ_{eff} of A₃₄₃ is 1.0 ns, τ_{eff} of A₃₄₄ is 1.0 ns, τ_{eff} of A₃₄₅ is 1.0 ns, τ_{eff} of A₃₄₆ is 1.0 ns, τ_{eff} of A₃₄₇ is 1.0 ns, τ_{eff} of A₃₄₈ is 1.0 ns, τ_{eff} of A₃₄₉ is 1.0 ns, τ_{eff} of A₃₅₀ is 1.0 ns, τ_{eff} of A₃₅₁ is 1.0 ns, τ_{eff} of A₃₅₂ is 1.0 ns, τ_{eff} of A₃₅₃ is 1.0 ns, τ_{eff} of A₃₅₄ is 1.0 ns, τ_{eff} of A₃₅₅ is 1.0 ns, τ_{eff} of A₃₅₆ is 1.0 ns, τ_{eff} of A₃₅₇ is 1.0 ns, τ_{eff} of A₃₅₈ is 1.0 ns, τ_{eff} of A₃₅₉ is 1.0 ns, τ_{eff} of A₃₆₀ is 1.0 ns, τ_{eff} of A₃₆₁ is 1.0 ns, τ_{eff} of A₃₆₂ is 1.0 ns, τ_{eff} of A₃₆₃ is 1.0 ns, τ_{eff} of A₃₆₄ is 1.0 ns, τ_{eff} of A₃₆₅ is 1.0 ns, τ_{eff} of A₃₆₆ is 1.0 ns, τ_{eff} of A₃₆₇ is 1.0 ns, τ_{eff} of A₃₆₈ is 1.0 ns, τ_{eff} of A₃₆₉ is 1.0 ns, τ_{eff} of A₃₇₀ is 1.0 ns, τ_{eff} of A₃₇₁ is 1.0 ns, τ_{eff} of A₃₇₂ is 1.0 ns, τ_{eff} of A₃₇₃ is 1.0 ns, τ_{eff} of A₃₇₄ is 1.0 ns, τ_{eff} of A₃₇₅ is 1.0 ns, τ_{eff} of A₃₇₆ is 1.0 ns, τ_{eff} of A₃₇₇ is 1.0 ns, τ_{eff} of A₃₇₈ is 1.0 ns, τ_{eff} of A₃₇₉ is 1.0 ns, τ_{eff} of A₃₈₀ is 1.0 ns, τ_{eff} of A₃₈₁ is 1.0 ns, τ_{eff} of A₃₈₂ is 1.0 ns, τ_{eff} of A₃₈₃ is 1.0 ns, τ_{eff} of A₃₈₄ is 1.0 ns, τ_{eff} of A₃₈₅ is 1.0 ns, τ_{eff} of A₃₈₆ is 1.0 ns, τ_{eff} of A₃₈₇ is 1.0 ns, τ_{eff} of A₃₈₈ is 1.0 ns, τ_{eff} of A₃₈₉ is 1.0 ns, τ_{eff} of A₃₉₀ is

Multimodal Scientific Data with Vector Databases

AI-ready vector databases provide the scalable vector storage and retrieval essential for developing and deploying massive foundation models that rely on vector representations of data.



Text



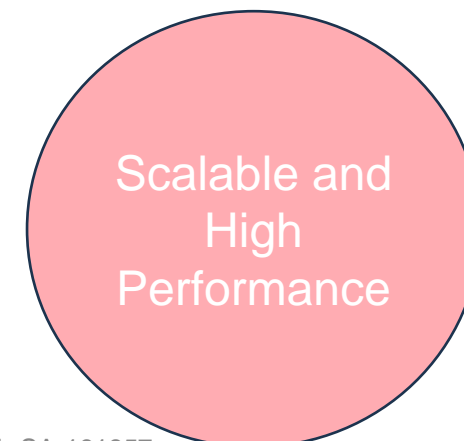
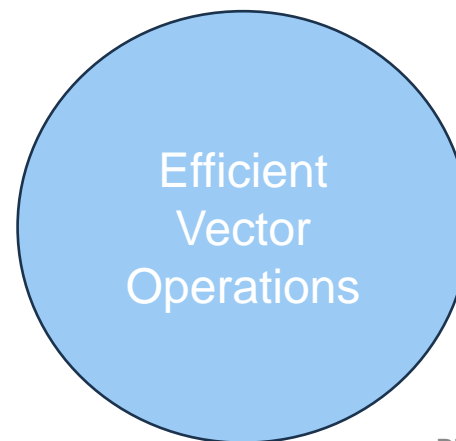
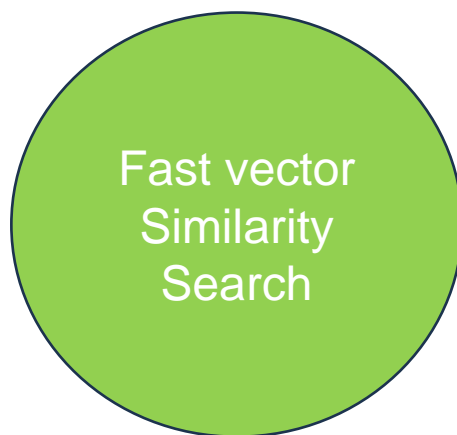
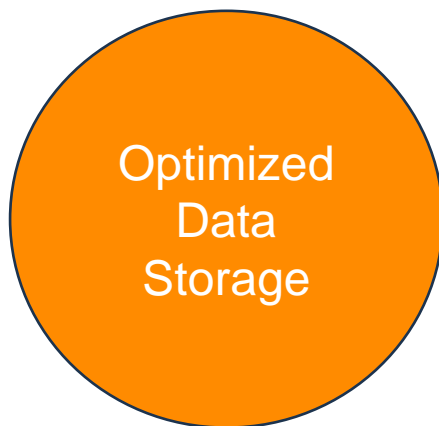
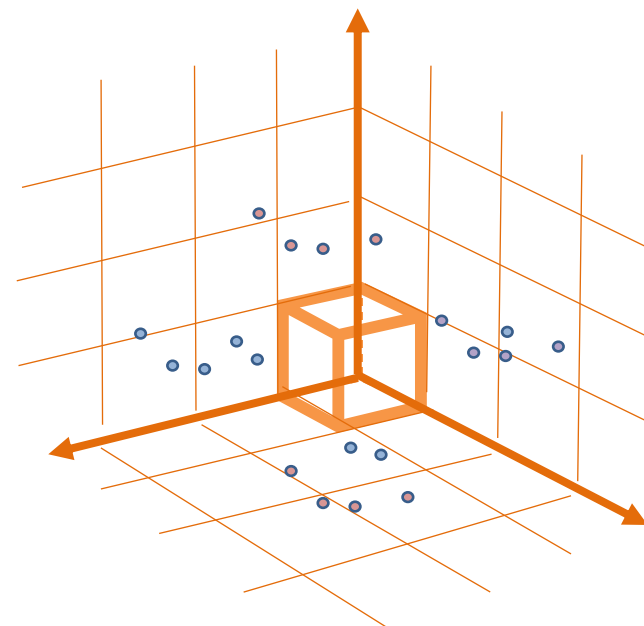
Vision



Tabular

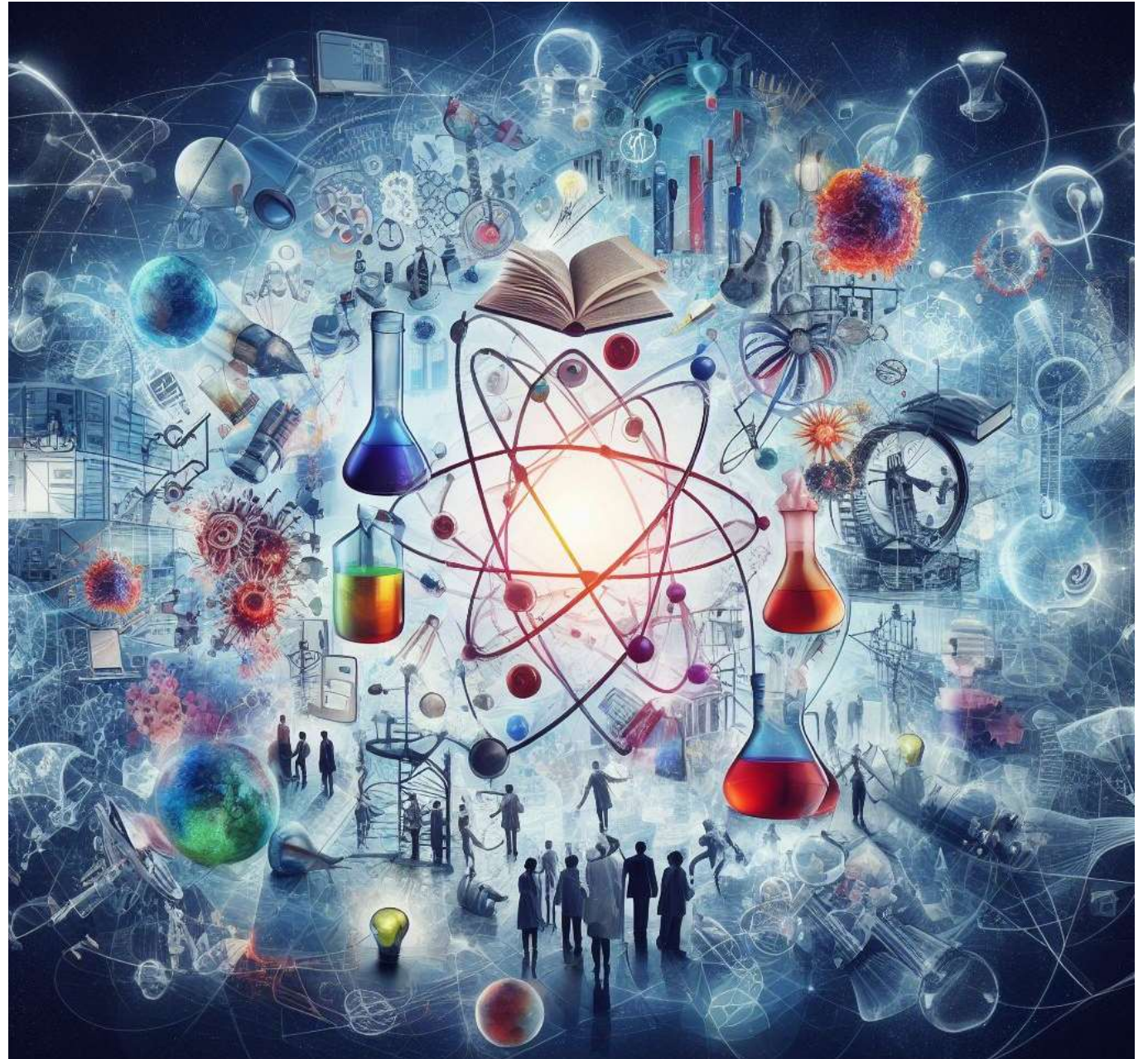


Graphs



ATLANTIC: Structure-Aware LLMs for Interdisciplinary Science

Developing LLMs to reason across multiple scientific disciplines



Cross-Disciplinarity Scientific Reasoning

Develop an integrated strategy to improve solar technology, drive down costs, model environmental benefits, and ensure inclusive deployment of solar energy across societies.



Physicist

investigating novel materials that can absorb and convert sunlight into electricity at higher efficiencies



Chemist

synthesizing and testing different compounds to find ones optimal for photon absorption.



Engineer

designing nanoscale architectures for solar cells to minimize electron-hole recombination losses.



Climate Scientist

modeling how increased adoption of solar energy could reduce greenhouse gas emissions from electricity production over time.



Economist

analyzing solar policies and incentives to understand how to further drive down costs and encourage widespread adoption.



Sociologist

studying how solar energy can be equitably distributed so all communities have access to clean electricity.

Cross-Document Scientific Reasoning

What is the name of the nuclear fuel used in the reactor?

Uranium Oxide (UO₂)

Uranium oxide (UO₂) is important from the technological point of view as a nuclear fuel. It has also been a challenging material for electronic structure calculations. On the one hand, being a nuclear fuel, understanding of thermal properties such as specific heat, thermal expansion and thermal conductivity in terms of structure at the atomistic/electronic scale is of great interest. On the other hand, complexity of the electronic structure makes material description very challenging. In the context of calculating material properties from the first principles, UO₂ has proven to be a non-trivial material. Conventional density functional theory (DFT) with local density approximation (LDA) and/or generalized gradient approximation (GGA), which has been quite successful in predicting the ground state structure and properties of a wide range of materials, has failed to give the proper description of UO₂.

Intra-Document Reasoning

Which country collaborated in the development of Fast Breeder Test reactor at the Indira Gandhi Center for Atomic Research?

Fast breeder reactors constitute the second stage of India's three-stage nuclear energy programme, for effective utilization of the country's limited reserves of natural uranium and exploitation of its large reserves of thorium. The Reactor Research Centre (renamed **Indira Gandhi Centre for Atomic Research** in 1985) was established at Kalpakkam, 80 km south of Madras (now Chennai), in 1971, with the mission to develop the technology of sodium cooled fast reactors.

Similar Topic
Fast Breeder Reactor

At the heart of the Reactor Research Centre was proposed a sodium cooled test reactor, named Fast Breeder Test Reactor, which would serve as a test bed for irradiation of fuels and materials and provide experience in large scale sodium handling and reactor operation. An agreement was signed with CEA (**France**) for transfer of the design of the Rapsodie reactor, training of personnel in Rapsodie and transfer of manufacturing technology of critical components.

Automated Reasoning Steps

France

The Reactor Research Centre is renamed to Indira Gandhi Center for Atomic Research

The Reactor Research Centre developed Fast Breeder Reactor

Fast Breeder Reactor is based on Rapsodie Reactor

Rapsodie Reactor is deployed at The French Alternative Energies and Atomic Energy Commission or CEA

CEA is in France

Cross-Documents Reasoning

Interdisciplinary Scientific Datasets

Content (Text) Representations



Interdisciplinary Scientific Datasets

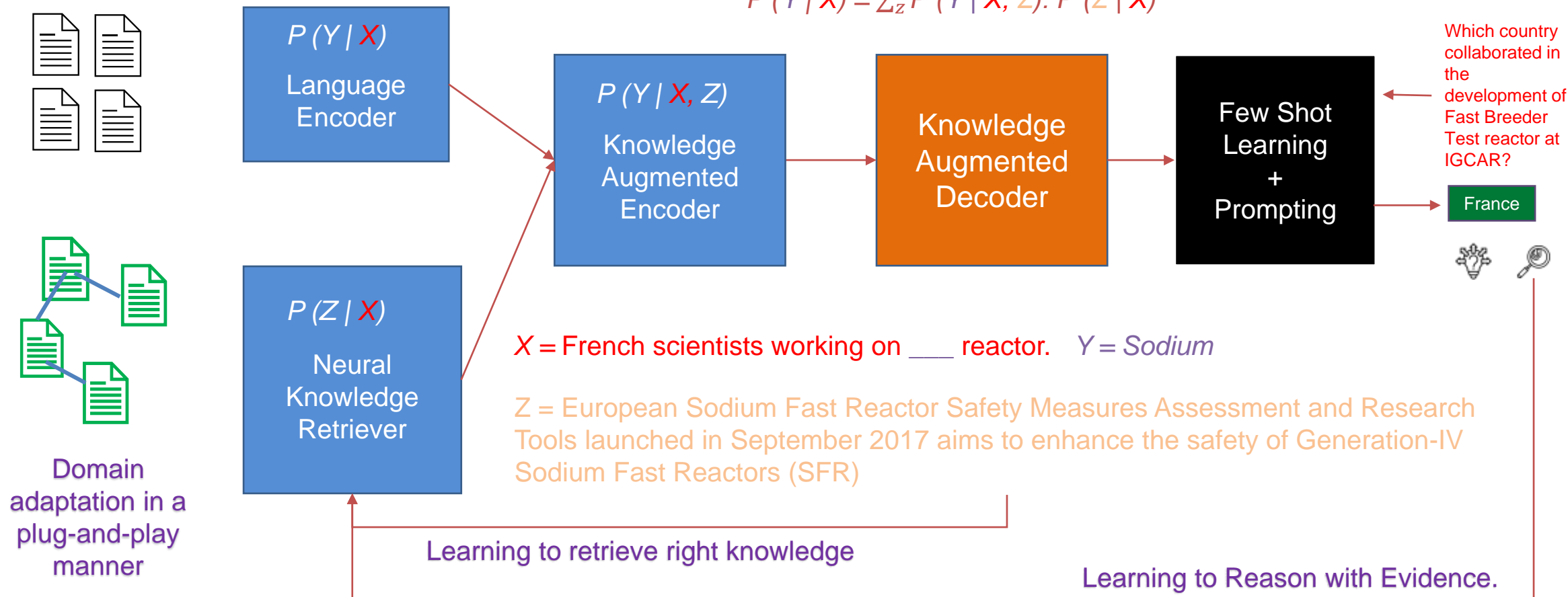
Content (Text) + Context (Graph) Representations



ATLANTIC Model Architecture

X = French scientists working on ____ reactor. Y = Sodium

$$P(Y|X) = \sum_z P(Y|X, Z) \cdot P(Z|X)$$



ATLANTIC Model Performance

Scientific Evidence Generation

How useful are the evidences generated from structural aware LLMs to justify model predictions in science tasks?

Atlas

"# Below is an input containing a title-abstract pair. Classify this input into one or more possible Field of Study categories [...]"

Input: ## Title: Characterization of N-Glycosylation Status of Micro-level Glycoprotein Using Solid Phase Extraction Based Techniques.
Abstract: A new strategy integrating the latest hydrophilic interaction chromatographic(HILIC) method with mass spectrometry technologies was developed for the enrichment and determination of intact glycopeptides and glycan structures from a glycoprotein. The average molecular weight, the ratio of glycosylation, the glycosite and the structures of glycoforms of ribonuclease B(RNaseB) were determined by MALDI-TOF-MS and ESI-LTQ-FT-MS. RNaseB contains 8.43% glycan and has a single glycosylation site at Asn34 with the glycan structure varied from five to nine mannose residues(Man5-9GlcNAc2). The strategy paves the way for systematic and confident analysis of complex glycoproteins with important biological or pharmaceutical functions.

Response: **Chemistry**

Evidence traces:

Passage 1: **[Geology]** Title: Paleomagnetic reversals in Miocene dikes, and tectonic evolution of the Crossman block, Mohave Mountains, Arizona

Passage 2: **[Social Science]** Title: Q&A with Photographer Paul Vecsei

Atlantic

"# Below is an input containing a title-abstract pair. Classify this input into one or more possible Field of Study categories [...]"

Input: ## Characterization of N-Glycosylation Status of Micro-level Glycoprotein Using Solid Phase Extraction Based Techniques ## Abstract: A new strategy integrating the latest hydrophilic interaction chromatographic(HILIC) method with mass spectrometry technologies was developed for the enrichment and determination of intact glycopeptides and glycan structures from a glycoprotein. The average molecular weight, the ratio of glycosylation, the glycosite and the structures of glycoforms of ribonuclease B(RNaseB) were determined by MALDI-TOF-MS and ESI-LTQ-FT-MS. RNaseB contains 8.43% glycan and has a single glycosylation site at Asn34 with the glycan structure varied from five to nine mannose residues(Man5-9GlcNAc2). The strategy paves the way for systematic and confident analysis of complex glycoproteins with important biological or pharmaceutical functions.

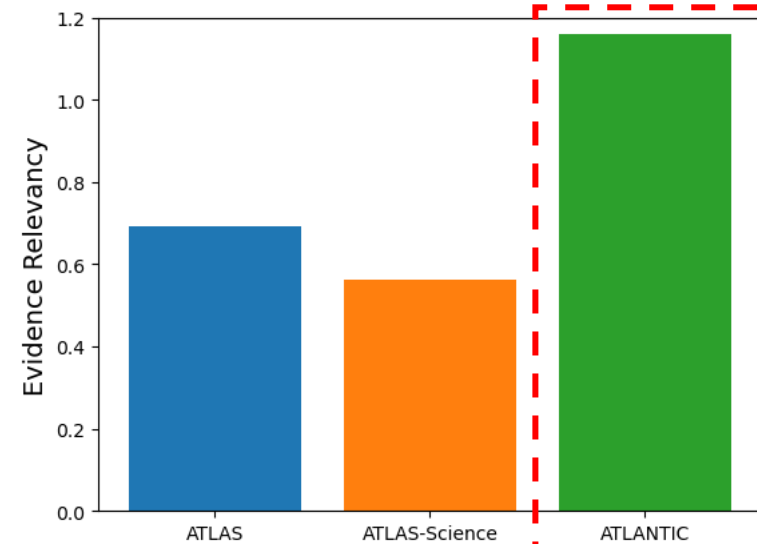
Response: **Chemistry**

Evidence traces:

Passage 1: **[Chemistry/Biology]** Title: Aryl hydrocarbon receptor in combination with Stat1 regulates LPS-induced inflammatory responses

Passage 2: **[Chemistry/Biology]** Title: Osteogenic protein-1 attenuates the inflammatory cytokine-induced NP cell senescence through regulating the ROS/NF-κB pathway.

Intra-disciplinary Scientific Reasoning
(Document Classification)



Structure aware LLMs retrieve relevant passages to justify model predictions better than text-only models

The query relevance metric calculates the semantic similarity of the extracted passages with the input query via their embeddings. Similarity scores are obtained via the dot product of the embeddings.

ATLANTIC Model Performance

Scientific Evidence Generation

How useful are the evidences generated from structural aware LLMs to justify model predictions in science tasks?

Atlas

"# question: Which of the following is the most likely benefit of periodic natural wildfires in some forest ecosystems? options: (A) The fires remove dead and decaying plant matter, reducing the risk of more intense and destructive fires. (B) The fires leach nutrients from the soil, preventing the germination of plants that might compete with native species. (C) The fires drive off herbivores whose plant-based diets reduce the amount of vegetation. (D) The fires dry out the soil and decrease the chance of flooding after heavy rains.

Response: A

Evidence traces:

Passage 1: [Art, History] Title: THE 'LAST BOLSHEVIK' - MARKER,C

Passage 2: [Maths, Engineering] Title: C Frame Transformations

Atlantic

"# question: Which of the following is the most likely benefit of periodic natural wildfires in some forest ecosystems? options: (A) The fires remove dead and decaying plant matter, reducing the risk of more intense and destructive fires. (B) The fires leach nutrients from the soil, preventing the germination of plants that might compete with native species. (C) The fires drive off herbivores whose plant-based diets reduce the amount of vegetation. (D) The fires dry out the soil and decrease the chance of flooding after heavy rains.

Response: A

Evidence traces:

Passage 1: [Geology] Title: Another Threat to Borneo's Rainforests?

Passage 2: [Geology] Title: The potential of secondary forests

Atlas

"# question: The Minoan economy may have been especially vulnerable to natural catastrophe because it relied heavily on : options: (A) agriculture (B) maritime trade (C) metalworking (D) slave labor answer

Response: B

Evidence traces:

Passage 1: [Computer science] Title: B Unicode Representation

Passage 2: [Maths, Engineering] Title: C Frame Transformations

Atlantic

"# question: The Minoan economy may have been especially vulnerable to natural catastrophe because it relied heavily on : options: (A) agriculture (B) maritime trade (C) metalworking (D) slave labor answer.

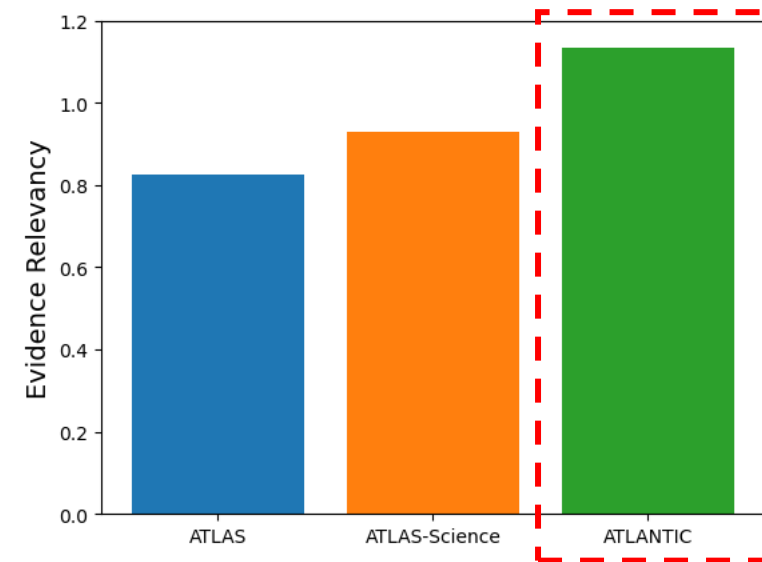
Response: B

Evidence traces:

Passage 1: [Economics] Title: Public debt and endogenous growth

Passage 2: [Economics] Title: Land and rent

Inter-disciplinary Scientific Reasoning
(Science Question and Answering)



Structure aware LLMs retrieve relevant passages to justify model predictions better than text-only models

The query relevance metric calculates the semantic similarity of the extracted passages with the input query via their embeddings. Similarity scores are obtained via the dot product of the embeddings.

ATLANTIC Model Performance

Faithfulness (Accuracy + Evidence Generation)

Does retrieving structural knowledge help to improve the overall model performance?

Yes

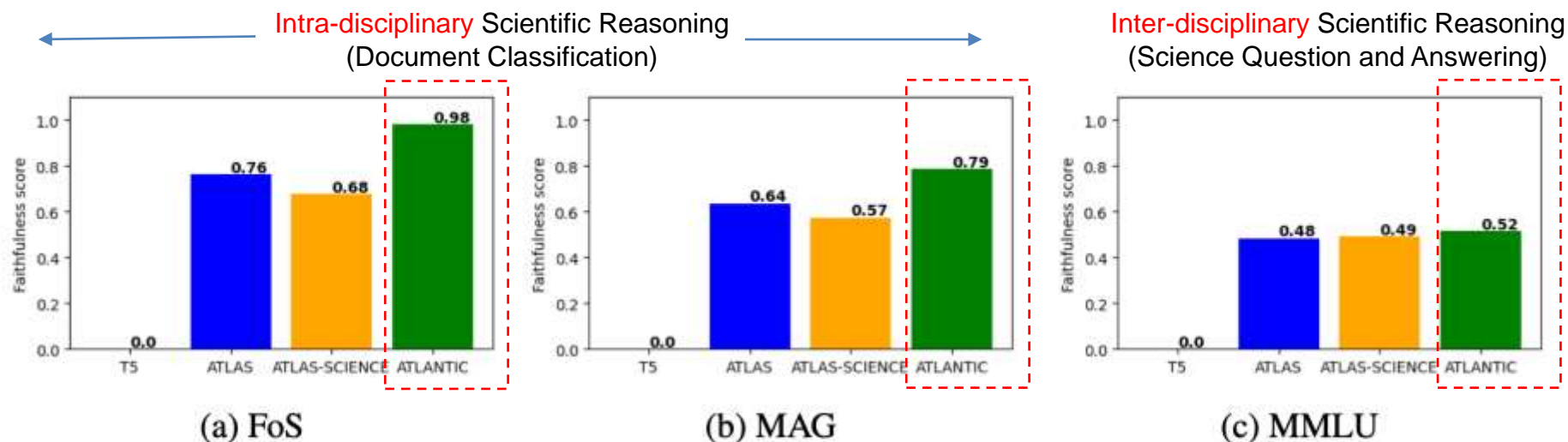


Figure 2: Faithfulness scores across FOS, MAG and MMLU benchmarks. Faithfulness score is the harmonic mean between the accuracy and relevance of the retrieved passages, which gives a holistic view on the trustworthiness of the model.

Faithfulness score is a measure combining generation accuracy and relevance score of the retrieved passages. T5, ATLAS, and ATLAS-Science are baseline models trained only with textual knowledge. T5 does not generate evidence such that the faithfulness score is zero.

ATLANTIC Model Demo

Human-AI Teaming

Text Based Reasoning

Which two isotopes are used in the majority of all neutron generators that are commercially available?

Text Query Log

Which two isotopes are used in the majority of all neutron generators that are commercially available?

tritium
Uncertainty: 40.0%

Source Documents

2006

D-D neutron generator development at LBNL.

...

There are three distinctive developments, which are discussed in this presentation, namely, multi-stage, accelerator-based axial neutron generator, high-output co-axial neutron generator and point source neutron generator. These generators ...

2003

Fission multipliers for D-D/D-T neutron generators

...

The high neutron yield makes this new generation of neutron generator suitable for a wide variety of applications for which those conventional neutron generators are unable to provide sufficient flux.

2000

Multipurpose neutron generators based on the radio frequency quadrupole linear

...

Since they do not use radioactive materials, licensing requirements are less stringent than for isotopic sources or tritium sealed tube generators.

1988

CANUTRON — A clean accelerator neutron generator

...

For applications requiring a lower intensity neutron source, a pulsed linac design with an average current of 1 mA has been studied. This option would reduce the operating cost by an additional 40%. Characteristics of both ...

1982

Technetium-99m generators--the available options.

...

Recent advances in generator technology indicate that neutron-activation-produced 99Mo may eventually replace the need for fission-produced 99Mo. The review mentions one method of achieving this goal.

1969

Thick target neutron yields and neutron spectra produced by 20 MeV helium-3

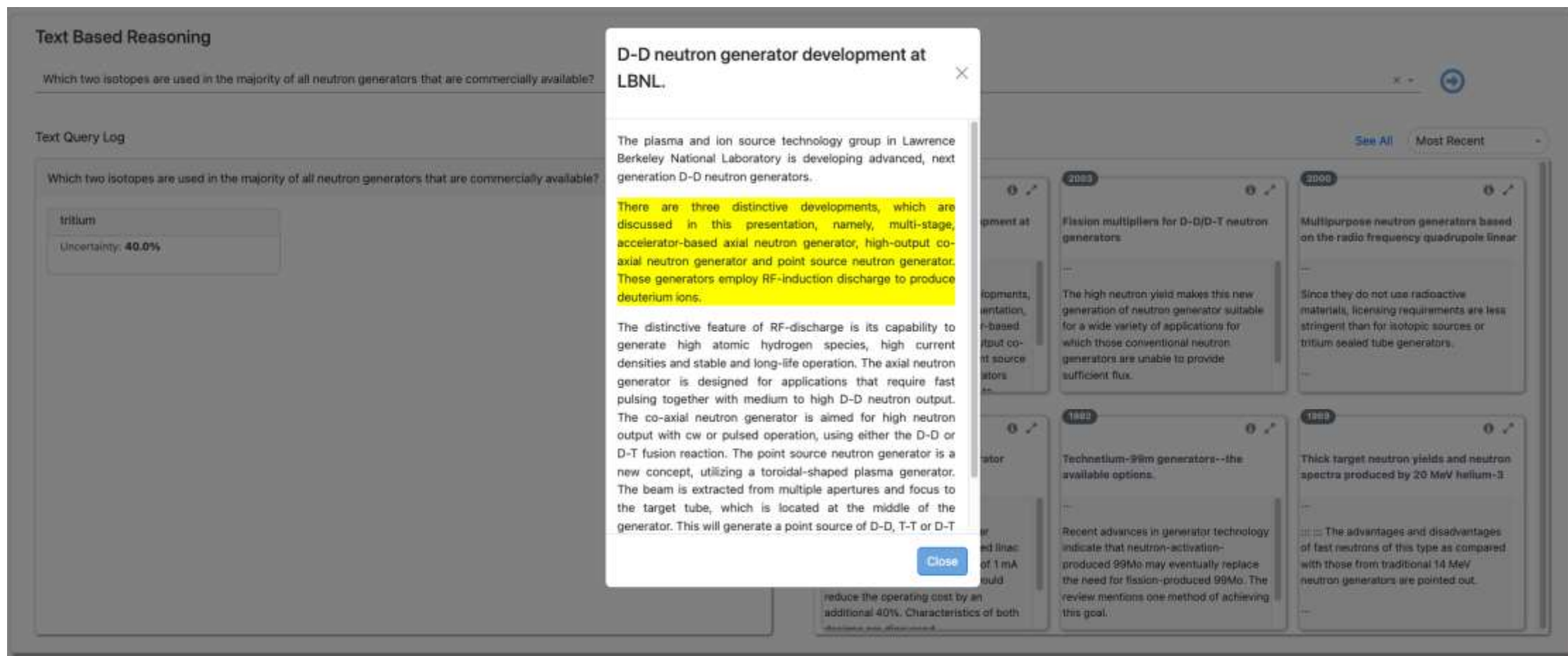
...

... The advantages and disadvantages of fast neutrons of this type as compared with those from traditional 14 MeV neutron generators are pointed out.

EXPERT 2.0 Human-AI Scientific Reasoning Engine <https://github.com/pnnl/EXPERT2>

ATLANTIC Model Demo

Human-AI Teaming



Text Based Reasoning

Which two isotopes are used in the majority of all neutron generators that are commercially available?

Text Query Log

Which two isotopes are used in the majority of all neutron generators that are commercially available?

tritium
Uncertainty: 40.0%

D-D neutron generator development at LBNL.

The plasma and ion source technology group in Lawrence Berkeley National Laboratory is developing advanced, next generation D-D neutron generators.

There are three distinctive developments, which are discussed in this presentation, namely, multi-stage, accelerator-based axial neutron generator, high-output co-axial neutron generator and point source neutron generator. These generators employ RF-induction discharge to produce deuterium ions.

The distinctive feature of RF-discharge is its capability to generate high atomic hydrogen species, high current densities and stable and long-life operation. The axial neutron generator is designed for applications that require fast pulsing together with medium to high D-D neutron output. The co-axial neutron generator is aimed for high neutron output with cw or pulsed operation, using either the D-D or D-T fusion reaction. The point source neutron generator is a new concept, utilizing a toroidal-shaped plasma generator. The beam is extracted from multiple apertures and focus to the target tube, which is located at the middle of the generator. This will generate a point source of D-D, T-T or D-T

2003
Fission multipliers for D-D/D-T neutron generators.

2006
Multipurpose neutron generators based on the radio frequency quadrupole linear

1992
Technetium-99m generators--the available options.

1989
Thick target neutron yields and neutron spectra produced by 20 MeV helium-3.

EXPERT 2.0 Human-AI Scientific Reasoning Engine <https://github.com/pnnl/EXPERT2>

SciTune: Multimodal Agents for Scientific Reasoning

“A llama in a science lab coat”

Developing multimodal agents to reason across multiple scientific disciplines



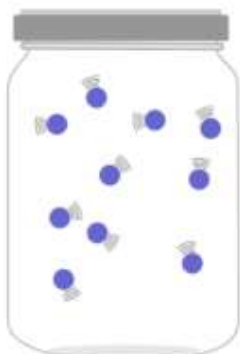
Scientific Multimodal Reasoning

Question

Compare the average kinetic energies of the particles in each sample. Which sample has the higher temperature?

Context

The diagrams below show two pure samples of gas in identical closed, rigid containers. Each colored ball represents one gas particle. Both samples have the same number of particles.



Sample A

Mass of each particle: 32 u

Average particle speed: 1,000 m/s



Sample B

Mass of each particle: 40 u

Average particle speed: 1,000 m/s

Choices

sample A

sample B

neither; the samples have the same temperature

Physics Example

Question

Which of the following could Kathleen's test show?

Context

People can use the engineering-design process to develop solutions to problems. One step in the process is testing if a potential solution meets the requirements of the design. The passage below describes how the engineering-design process was used to test a solution to a problem. Read the passage. Then answer the question below. People with diabetes sometimes take a medicine made from insulin. Insulin can be made by a special type of bacteria. Kathleen was a bioengineer who wanted to increase the amount of insulin that the bacteria produced by 20%. She read that giving the bacteria more nutrients could affect the amount of insulin they produced. So, Kathleen gave extra nutrients to some of the bacteria. Then, she measured how much insulin those bacteria produced compared to bacteria that did not get extra nutrients. Figure: studying bacteria in a laboratory.



Choices

whether she added enough nutrients to help the bacteria produce 20% more insulin

whether producing more insulin would help the bacteria grow faster

whether different types of bacteria would need different nutrients to produce insulin

Science and Eng. Practices Example

Interdisciplinary Scientific Datasets

Multimodal Content (Text + Images) Representations



AI-Ready Vector Databases

SciTune Multistage Concept and Task Alignment

How to make an LLM that can see and reason in the scientific multimodal world?

How to chain vision experts with LLMs?

How to learn scientific image representations?

Which human-generated scientific data modalities support to align LLMs with scientific concepts?

How to generate visual instructions?



How to perform scientific multimodal reasoning with visual grounded QA and text-driven explanations?

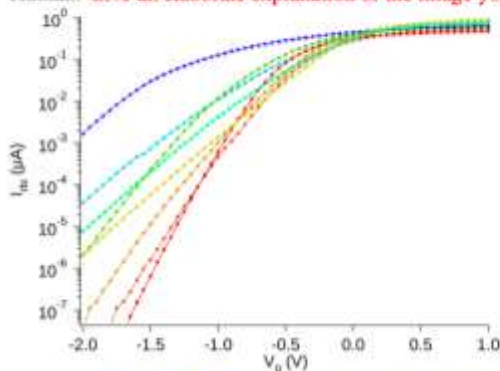
How to evaluate the scientific multimodal reasoning?

SciTune Multimodal Token Generation

SciTune Multimodal Instruction Examples. We distinguish the system message s_d , natural language instruction s_i , scientific figure type s_t , caption s_c , OCR s_o and paragraph mentions s_m .

A chat between a curious human and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the human's questions.

Human: Give an elaborate explanation of the image you see.

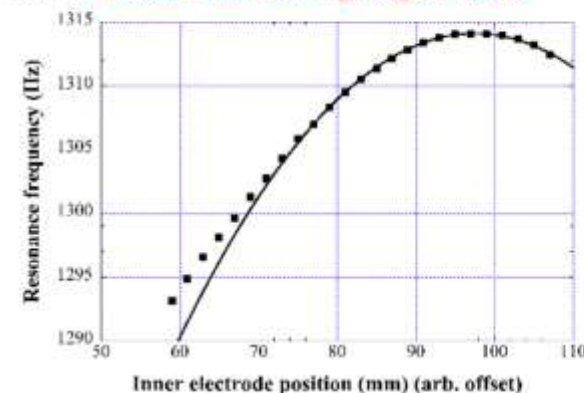


Assistant: **Graph Plot** Room temperature characteristics (drain-source current versus gate voltage) for a set of nominally identical samples from the same wafer. Sample to sample variations are very important, especially below the threshold voltage (-0.5 V for the best device, in red). In the most extreme case (blue curve) the many decades of excess current compared to the best device are attributed to direct transport through individual arsenic dopants which have diffused into the channel after doping and annealing steps. The source-drain voltage is 10mV for these data.

FIG. 6: figure S1: Room temperature characteristics (drain-source current versus gate voltage) for a set of nominally identical samples from the same wafer.

A chat between a curious human and an artificial intelligence assistant. The assistant gives helpful, detailed, and polite answers to the human's questions.

Human: Describe the following image in detail.



Assistant: **Graph Plot** Resonance frequency of the (1, 0) mode of the $r_1 = 4$ cm cavity filled with 1 atm of N2 as a function of the vertical position of the central electrode.

Inner electrode position, (arb , offset)

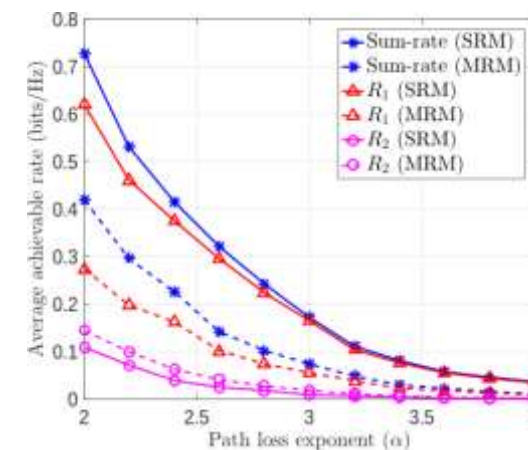
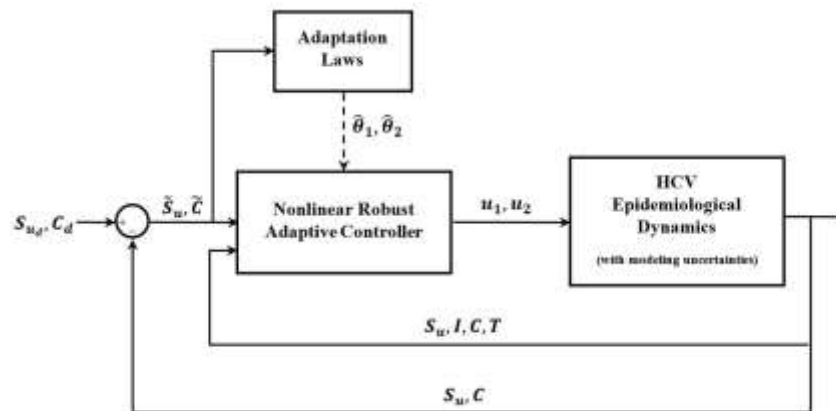
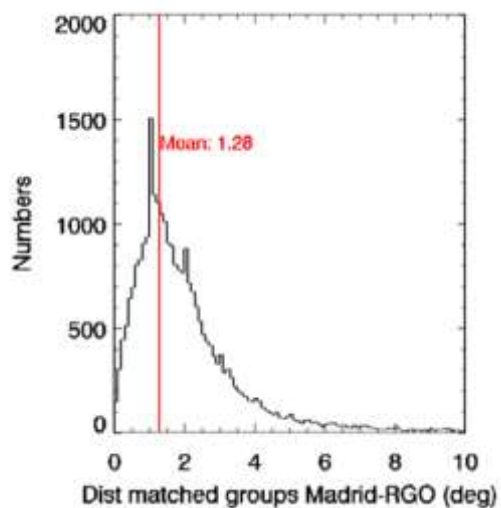
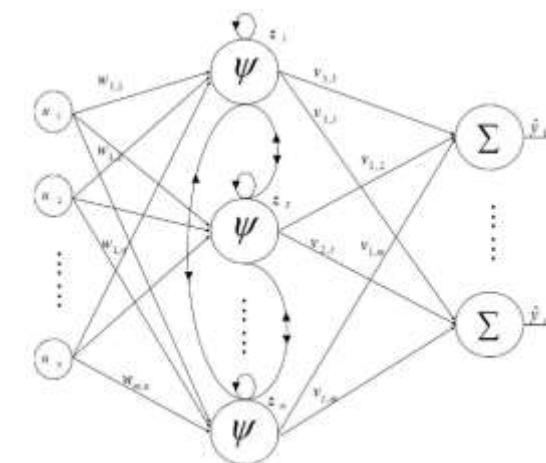
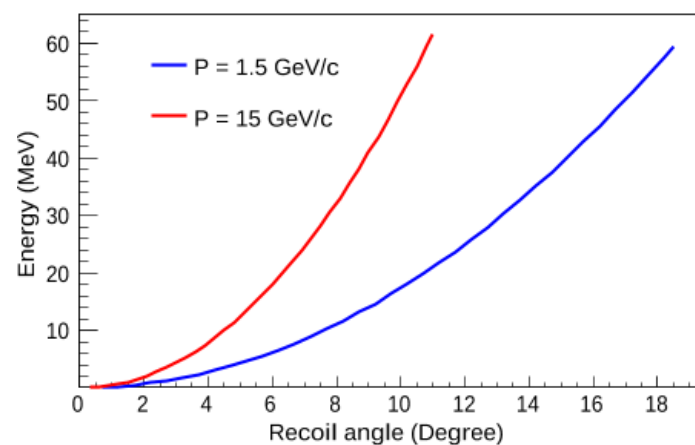
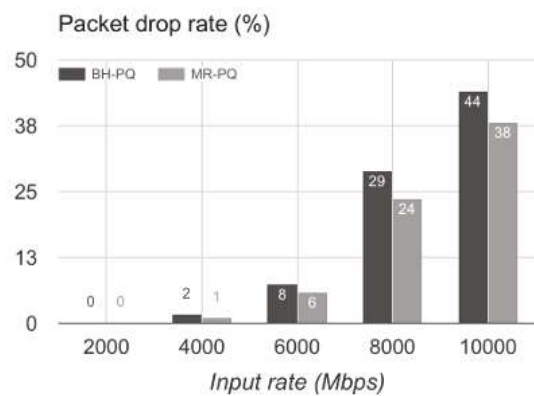
In Fig. 10 we have an example of this behavior: the resonance frequency shows a maximum which corresponds to a minimum in the eccentricity of the electrode [39].

$$p(x) = \prod_{j=1}^n p(s_{T>j} | s_V, s_I, s_{T<j})$$

SciTune Performance

Scientific Image Captioning

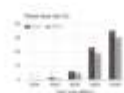
Testing the Scientific Concept Alignment Stage Pretraining



SciTune Performance

Scientific Image Captioning

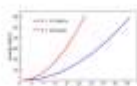
Table 8: A Sample of Generated Captions. We highlight the gold standard caption in red, and generated captions from the BLIP (Li et al., 2022) model in gray. LLaMA-SciTune model first generates the figure types followed with the captions colored in blue.



Packet drop rate

a chart of a bar chart with a number of different items

Bar Chart Packet drop rate for each method.



The kinetic energy of the recoil protons as a function of the recoil angle at beam momenta $P=1.5$ and 15 GeV/c, blue and red, respectively.

a plot of a curve with a blue line and a red line.

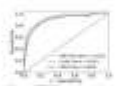
Graph Plot The angular distribution of the electron recoil spectrum in the 1.5 GeV/c and 15 GeV/c electron beams.



Artificial neural network structure.

a diagram of a network with several different paths.

Node Diagram The generative neural network.



ROC curves of cIBP-VAE in comparison to alternative models on the clinical ECG data set.

a plot of the average and average time of a cell phone.

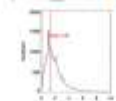
Graph Plot ROC curves of c-VAE, CNN, and c-VAE+CNN on the cerebellar atrophy dataset.



Functional architecture of the developed prototype.

a diagram of a camera and a person on a phone.

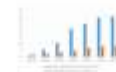
Node Diagram An overview of the system architecture of the proposed method.



Distance between matched groups in Madrid and RGO catalogs (bins of 0.1 degrees). The red line represents the mean value.

a plot of a line of data with a red line and a white line.

Graph Plot Distance correlation between groups matched by Madrid RGO.



Comparison of penetration rate of mobile broadband subscribers with that of fixed broadband subscribers.

a chart of the number of people who are using the internet.

Bar Chart The average rate of mobile broadband subscribers and fixed broadband subscribers for each quarter (in thousands).

Table 1: Accuracy of Generating the Figure Types. We also report the zero-shot figure type classification performance of the CLIP (Radford et al., 2021) model.

Figure Type	CLIP	LLaMA-SciTune
Graph Plot	54.07	93.48
Scatterplot	53.48	79.06
Node Diagram	91.02	98.71
Equation	65.71	94.28
Bar Chart	28.94	84.21
All	58.68	92.42

Table 2: Evaluation of Generated Figure Captions

Model	BLEU	ROUGE
BLIP (Li et al., 2022)	0.02 ± 0.02	0.10 ± 0.07
LLaMA-SciTune	0.05 ± 0.03	0.13 ± 0.08

SciTune Performance

ScienceQA Multimodal Reasoning

3 subject areas, 26 topics, 127 categories, and 379 skills

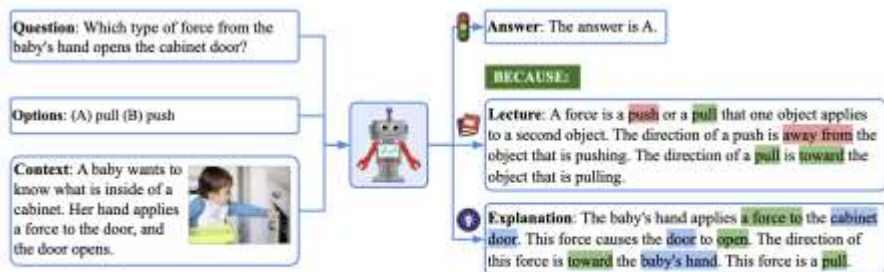
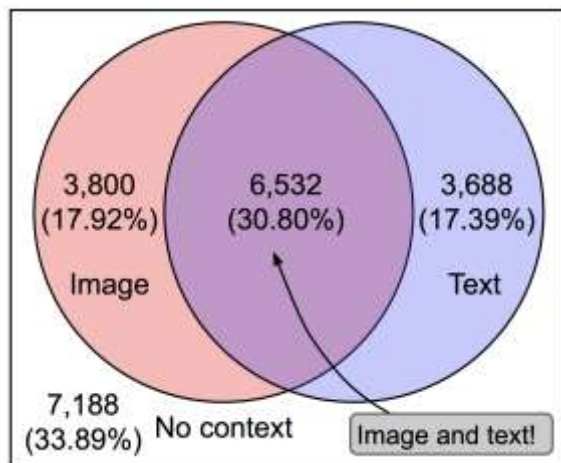


Figure 1: We construct the SCIENCEQA dataset where a data example consists of multimodal question answering information and the grounded lecture and explanation. We study if QA models can generate a reasonable explanation to reveal the chain-of-thought reasoning.



Statistic	Number
Total questions	21,208
Questions with text context	10,220 (48.2%)
Questions with image context	10,332 (48.7%)
* Image of natural format	≈2,960 (14.0%)
* Image of diagram format	≈7,372 (34.8%)
Questions with both contexts	6,532 (30.8%)
Questions without any context	7,188 (33.9%)
Questions with a lecture	17,798 (83.9%)
Questions with an explanation	19,202 (90.5%)
Different questions	9,122
Different lectures	261
Topic classes	26
Category classes	127
Skill classes	379
Average question length	12.11
Average choice length	4.40
Average lecture length	125.06
Average explanation length	47.66

Table 1: Main statistics in SCIENCEQA.



Figure 2: Question distribution in SCIENCEQA.

Grades	Number	Percent
Grade 1	95	0.45%
Grade 2	1,678	7.91%
Grade 3	3,032	14.3%
Grade 4	3,544	16.71%
Grade 5	3,086	14.55%
Grade 6	2,450	11.55%
Grade 7	2,749	12.96%
Grade 8	2,546	12.0%
Grade 9	491	2.32%
Grade 10	558	2.63%
Grade 11	539	2.54%
Grade 12	440	2.07%

(a) Grade distribution statistics.



(b) Grade distribution visualization.

ScienceQA is the SOTA Science-Focused Multimodal Reasoning Benchmark

SciTune Performance

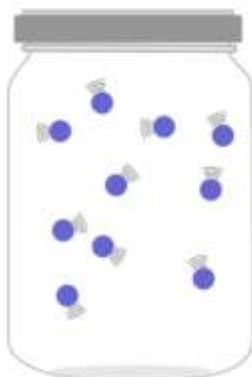
Natural Sciences - Physics

Question

Compare the average kinetic energies of the particles in each sample. Which sample has the higher temperature?

Context

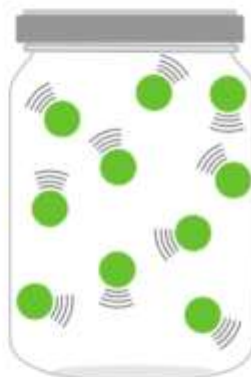
The diagrams below show two pure samples of gas in identical closed, rigid containers. Each colored ball represents one gas particle. Both samples have the same number of particles.



Sample A

Mass of each particle: 32 u

Average particle speed: 1,000 m/s



Sample B

Mass of each particle: 40 u

Average particle speed: 1,000 m/s

Choices

sample A

sample B

neither; the samples have the same temperature

Answer

sample B

Lecture

The temperature of a substance depends on the average kinetic energy of the particles in the substance. The higher the average kinetic energy of the particles, the higher the temperature of the substance. The kinetic energy of a particle is determined by its mass and speed. For a pure substance, the greater the mass of each particle in the substance and the higher the average speed of the particles, the higher their average kinetic energy.

Solution

The particles in both samples have the same average speed, but each particle in sample B has more mass than each particle in sample A. So, the particles in sample B have a higher average kinetic energy than the particles in sample A. Because the particles in sample B have the higher average kinetic energy, sample B must have the higher temperature.

Prediction



Assistant: LECTURE: The temperature of a substance depends on the average kinetic energy of the particles in the substance. The higher the average kinetic energy of the particles, the higher the temperature of the substance. The kinetic energy of a particle is determined by its mass and speed. For a pure substance, the greater the mass of each particle in the substance and the higher the average speed of the particles, the higher their average kinetic energy. SOLUTION: The particles in both samples have the same average speed, but each particle in sample B has more mass than each particle in sample A. So, the particles in sample B have a higher average kinetic energy than the particles in sample A. Because the particles in sample B have the higher average kinetic energy, sample B must have the higher temperature. The answer is B.

Accuracy Label

correct

SciTune Performance

Natural Sciences – Science and Eng. Practices

Question

Which of the following could Kathleen's test show?

Context

People can use the engineering-design process to develop solutions to problems. One step in the process is testing if a potential solution meets the requirements of the design. The passage below describes how the engineering-design process was used to test a solution to a problem. Read the passage. Then answer the question below. People with diabetes sometimes take a medicine made from insulin. Insulin can be made by a special type of bacteria. Kathleen was a bioengineer who wanted to increase the amount of insulin that the bacteria produced by 20%. She read that giving the bacteria more nutrients could affect the amount of insulin they produced. So, Kathleen gave extra nutrients to some of the bacteria. Then, she measured how much insulin those bacteria produced compared to bacteria that did not get extra nutrients. Figure: studying bacteria in a laboratory.



Choices

whether she added enough nutrients to help the bacteria produce 20% more insulin

whether producing more insulin would help the bacteria grow faster

whether different types of bacteria would need different nutrients to produce insulin

Answer

whether she added enough nutrients to help the bacteria produce 20% more insulin

Lecture

People can use the engineering-design process to develop solutions to problems. One step in the process is testing if a potential solution meets the requirements of the design. How can you determine what a test can show? You need to figure out what was tested and what was measured. Imagine an engineer needs to design a bridge for a windy location. She wants to make sure the bridge will not move too much in high wind. So, she builds a smaller prototype, or model, of a bridge. Then, she exposes the prototype to high winds and measures how much the bridge moves. First, identify what was tested. A test can examine one design, or it may compare multiple prototypes to each other. In the test described above, the engineer tested a prototype of a bridge in high wind. Then, identify what the test measured. One of the criteria for the bridge was that it not move too much in high winds. The test measured how much the prototype bridge moved. Tests can show how well one or more designs meet the criteria. The test described above can show whether the bridge would move too much in high winds.

Solution

Prediction

Assistant: LECTURE: People can use the engineering-design process to develop solutions to problems. One step in the process is testing if a potential solution meets the requirements of the design. How can you determine what a test can show? You need to figure out what was tested and what was measured. Imagine an engineer needs to design a bridge for a windy location. She wants to make sure the bridge will not move too much in high wind. So, she builds a smaller prototype, or model, of a bridge. Then, she exposes the prototype to high winds and measures how much the bridge moves. First, identify what was tested. A test can examine one design, or it may compare multiple prototypes to each other. In the test described above, the engineer tested a prototype of a bridge in high wind. Then, identify what the test measured. One of the criteria for the bridge was that it not move too much in high winds. The test measured how much the prototype bridge moved. Tests can show how well one or more designs meet the criteria. The test described above can show whether the bridge would move too much in high winds. The answer is A.

Accuracy Label

correct

SciTune Performance

ScienceQA Performance (Accuracy)

LLaMA-SciTune (13B) model outperforms the human performance on average.

LLaMA-SciTune models pretrained with additional scientific modalities such as caption, figure type, OCR, and figure mentions perform better than the model pretrained only with captions.

There is a significant performance advantage of the models trained with larger language decoder model (13B) compared to the relatively smaller model (7B).

Table 3: Results (accuracy %) on ScienceQA dataset. Question classes: NAT = natural science, SOC = social science, LAN = language science, TXT = text context, IMG = image context, NO = no context, G1-6 = grades 1-6, G7-12 = grades 7-12. We present two variants, *LLaMA-SciTune (C)* and *LLaMA-SciTune (CTOM)*. Acronyms inside the parenthesis represent the input modalities used in the SciTune instruction template. E.g., Caption, Figure Type, OCR, and Figure Mentions. We use the notation ♠ to denote the models finetuned with GPT-3.5/4 synthetic instructions, or use GPT-3.5/4 for any support during the inference time. We bold the accuracy values that are greater than what humans achieved. For additional baseline results, please refer the public ScienceQA leaderboard²

Method	#Params	Avg	NAT	SOC	LAN	TXT	IMG	NO	G1-6	G7-12
Random Chance	-	39.83	40.28	46.13	29.25	47.45	40.08	33.66	39.35	40.67
Human Average	-	88.40	90.23	84.97	87.48	89.60	87.50	88.10	91.59	82.42
UnifiedQA	223M	70.12	68.16	69.18	74.91	63.78	61.38	77.84	72.98	65.00
UnifiedQA (CoT)	223M	74.11	71.00	76.04	78.91	66.42	66.53	81.81	77.06	68.82
♠ GPT-3 (Zero Shot)	175B	74.04	75.04	66.59	78.00	74.24	65.74	79.58	76.36	69.87
♠ GPT-3 (CoT) (ALE)	175B	75.17	75.44	70.87	78.09	74.68	67.43	79.93	78.23	69.68
♠ ChatGPT CoT	175B+	78.31	78.82	70.98	83.18	77.37	67.92	86.13	80.72	74.03
♠ GPT-4 CoT	1T+	83.99	85.48	72.44	90.27	82.65	71.49	92.89	86.66	79.04
Multimodal-CoT	223M	84.91	87.52	77.17	85.82	87.88	82.90	86.83	84.65	85.37
Multimodal-CoT	770M	91.68	95.91	82.00	90.82	95.26	88.80	92.89	92.44	90.31
♠ LLaMA-Adapter	13B	85.19	84.37	88.30	84.36	83.72	80.32	86.90	85.83	84.05
♠ LLaVa	13B	90.92	90.36	95.95	88.00	89.49	88.00	90.66	90.93	90.90
♠ LLaVa + GPT-4 (judge)	13B	92.53	91.56	96.74	91.09	90.62	88.99	93.52	92.73	92.16
♠ Chameleon (ChatGPT)	175B+	79.93	81.62	70.64	84.00	79.77	70.80	86.62	81.86	76.53
♠ Chameleon (GPT-4)	1T+	86.54	89.83	74.13	89.82	88.27	77.64	92.13	88.03	83.72
LLaMA-SciTune (C)	7B	85.61	84.36	92.23	82.81	89.56	81.26	88.29	81.28	86.03
LLaMA-SciTune (CTOM)	7B	86.11	84.50	94.15	82.91	88.35	83.64	88.74	85.05	85.60
LLaMA-SciTune (CTOM)	13B	90.03	89.30	95.61	87.00	93.08	86.67	91.75	84.37	91.30

SciTune Performance

ScienceQA Performance (Evidence)

Did the model accurately explain the reasoning that supports the answer?

Table 4: Few-shot performance analysis. We report the number of times lectures seen during the training in frequency, and the number of test questions with the lecture.

Frequency	#Questions	Accuracy (7B)	Accuracy (13B)
5	36	75.00	83.33
10	125	81.60	85.60
25	412	80.34	85.92
50	1140	81.05	86.14

The model generates the answer more accurately when it acquires the relevant background knowledge during training

Table 5: Evaluation of generated lectures and solutions.

	7B Model		13B Model	
	BLEU	ROUGE	BLEU	ROUGE
All answers				
Lecture	0.763	0.778	0.854	0.868
Solution	0.791	0.838	0.872	0.921
Correct answers				
Lecture	0.765	0.780	0.847	0.861
Solution	0.829	0.873	0.893	0.937
Incorrect answers				
Lecture	0.751	0.767	0.909	0.924
Solution	0.565	0.631	0.694	0.778

The model generates the solution text with higher accuracy than the lecture text.

Key Takeaways

Advancing Interdisciplinary Science with Foundation Models

Synthesize

Synthesis of
knowledge from
diverse scientific
disciplines

Adoption

Rapidly adapt to
new tasks with
minimal labeled
data

Scalability

Learning complex
concepts spanning
disciplinary boundaries
from massive datasets.

Accessibility

Higher level analysis,
hypothesis development and
knowledge integration
across disciplines

Foundation Models for Science Papers

- Munikoti, S., Acharya, A., Wagle, S., Horawalavithana, S., **ATLANTIC: Structure-Aware Retrieval-Augmented Language Model for Interdisciplinary Science** (preprint available upon request)
- Acharya, A., Munikoti, S., Hellinger, A., Smith, S., Wagle, S., Horawalavithana, S., **NuclearQA: A Human-Made Benchmark for Language Models for the Nuclear Domain** (2023) arXiv Preprint arXiv:2310. 10920.
- Munikoti, S., Acharya, A., Wagle, S., Horawalavithana, S., **Evaluating the Effectiveness of Retrieval-Augmented Large Language Models in Scientific Document Reasoning** (preprint available upon request)
- Wagle, S., Munikoti, S., Acharya, A., Horawalavithana, S., **Empirical evaluation of uncertainty quantification in retrieval-augmented language models for science** (preprint available upon request)
- Horawalavithana, S., Munikoti, S., Stewart, I., & Kvinge, H. (2023). **SCITUNE: Aligning Large Language Models with Scientific Multimodal Instructions**. arXiv preprint arXiv:2307.01139.
- Dollar, O. W., Horawalavithana, S., Vasquez, S., Pfaendtner, W. J., & Volkova, S. (2022). **MolJET: Multimodal Joint Embedding Transformer for Conditional de novo Molecular Design and Multi-Property Optimization**. <https://openreview.net/pdf?id=7UudBVslrr>
- Horawalavithana, S., Ayton, E., Sharma, S., Howland, S., Subramanian, M., Vasquez, S., ... & Volkova, S. (2022, May). **Foundation models of scientific knowledge for chemistry: Opportunities, challenges and lessons learned**. In Proceedings of BigScience Episode# 5--Workshop on Challenges & Perspectives in Creating Large Language Models (pp. 160-172).

Foundation Models to Advance Interdisciplinary Science

Sameera Horawalavithana, Ph.D.

yasanka.horawalavithana@pnnl.gov

